Multiclass Classification

Text Classification

A Cancer Conundrum: Too Many Drug Trials, Too Few Patients
Breakthroughs in immunotherapy and a rush to develop profitable new treatments have brought a crush of clinical trials scrambling for patients.

Yankees and Mets Are on Opposite Tracks This Subway Series
As they meet for a four-game series, the Yankees are playing for a postseason spot, and the most the Mets can hope for is to play spoiler.

20 Newsgroups, Reuters, Yahoo! Answers, ...

Entailment

A soccer game with multiple males playing.
ENTAILS
Some men are playing a sport.

A black race car starts up in front of a crowd of people.

A man is driving down a lonely road

A smiling costumed woman is holding an umbrella.
NEUTRAL
A happy woman in a fairy costume holds an umbrella.

Entity Linking

Although he originally won the event, the United States Anti-Doping Agency announced in August 2012 that they had disqualified Armstrong from his seven consecutive Tour de France wins from 1999–2005.

Lance Edward Armstrong is an American former professional road cyclist

Armstrong County is a county in Pennsylvania...

4,500,000 classes (all articles in Wikipedia)

Needs a very different structure for $f(x,y)$ to accommodate so many classes

Bowman et al. (2015)
Authorship Attribution

- Statistical methods date back to 1930s and 1940s
- Based on handcrafted heuristics like stopword frequencies
- Early work: Shakespeare’s plays, Federalist papers (Hamilton vs. Madison)
- Twitter: given a bunch of tweets, can we figure out who wrote them?
  - Schwartz et al. EMNLP 2013: 500M tweets, take 1000 users with at least 1000 tweets each
  - Task: given a held-out tweet by one of the 1000 authors, who wrote it?

- SVM with character 4-grams, word 2-grams through 5-grams
- 1000 authors, 200 tweets per author => 30% accuracy
- 50 authors, 200 tweets per author => 71.2% accuracy

---

<table>
<thead>
<tr>
<th>Signature Type</th>
<th>10%-signature</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character n-grams</td>
<td>&quot; ~ &quot;</td>
<td>REF oh ok 😂 Glad you found it!</td>
</tr>
<tr>
<td></td>
<td>&quot; ~ &quot;</td>
<td>Hope everyone is having a good afternoon 😄</td>
</tr>
<tr>
<td></td>
<td>&quot; ~ &quot;</td>
<td>REF Smirnoff lol keeping the goose in the freezer 😆</td>
</tr>
<tr>
<td></td>
<td>'yew '</td>
<td>gurl yew serving me tea nooch</td>
</tr>
<tr>
<td></td>
<td>'yew '</td>
<td>REF about wen yew and Ronnie see each other</td>
</tr>
<tr>
<td></td>
<td>'yew '</td>
<td>REF lol so yew goin to check out timi’s tonight huh??</td>
</tr>
</tbody>
</table>

---

Schwartz et al. (2013)